Randomness-Enhanced Expressivity of Quantum Neural Networks

Yadong Wu,[1,2,3] Juan Yao,[4,5,6] Pengfei Zhang,[1,3,*] and Xiaopeng Li[1,2,3,7,8,†]
[1]Department of Physics, Fudan University, Shanghai 200438, China
[2]State Key Laboratory of Surface Physics, Key Laboratory of Micro and Nano Photonic Structures (MOE), Institute for Nanoelectronic Devices and Quantum Computing, Fudan University, Shanghai 200438, China
[3]Shanghai Qi Zhi Institute, AI Tower, Xuhui District, Shanghai 200232, China
[4]Shenzhen Institute for Quantum Science and Engineering, Southern University of Science and Technology, Shenzhen, 518055 Guangdong, China
[5]International Quantum Academy, Shenzhen, 518048 Guangdong, China
[6]Guangdong Provincial Key Laboratory of Quantum Science and Engineering, Southern University of Science and Technology, Shenzhen, 518055 Guangdong, China
[7]Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China
[8]Shanghai Research Center for Quantum Sciences, Shanghai 201315, China

(Received 1 September 2023; revised 6 November 2023; accepted 6 December 2023; published 5 January 2024)

As a hybrid of artificial intelligence and quantum computing, quantum neural networks (QNNs) have gained significant attention as a promising application on near-term, noisy intermediate-scale quantum devices. Conventional QNNs are described by parametrized quantum circuits, which perform unitary operations and measurements on quantum states. In this Letter, we propose a novel approach to enhance the expressivity of QNNs by incorporating randomness into quantum circuits. Specifically, we introduce a random layer, which contains single-qubit gates sampled from a trainable ensemble pooling. The prediction of QNN is then represented by an ensemble average over a classical function of measurement outcomes. We prove that our approach can accurately approximate arbitrary target operators using Uhlmann's theorem for majorization, which enables observable learning. Our proposal is demonstrated with extensive numerical experiments, including observable learning, Rényi entropy measurement, and image recognition. We find the expressivity of QNNs is enhanced by introducing randomness for multiple learning tasks, which could have broad application in quantum machine learning.

DOI: 10.1103/PhysRevLett.132.010602

Introduction.—In recent years, significant breakthroughs have been made in the field of artificial intelligence. Among various machine learning algorithms, neural networks have played a vital role, thanks to their universal expressivity for deep architectures. As a quantum generalization of neural networks, quantum neural networks (QNNs) have been proposed based on parametrized quantum circuits. QNNs use quantum states instead of classical numbers as inputs [1–4]. However, the evolution of the input quantum states is constrained to be unitary, which limits the expressivity of QNNs. For physical observables, which are linear functions of the input quantum states or density matrices, QNNs can achieve high accuracy only if the target operator shares the same eigenvalues with the measurement operator. For a general situation, it requires introducing auxiliary qubits, as proposed in [5]. To express nonlinear functions of the input density matrices, such as purities, traditional approaches introduce multiple replicas, which is unfavorable on near-term, noisy intermediate-scale quantum (NISQ) devices with a limited number of logical qubits. Previous studies have also reported moderate accuracy for more general machine learning tasks, including image recognition [6–10].

In this Letter, we propose a universal scheme to overcome the expressivity obstacle without the need for additional replicas. Our main inspiration comes from the recent development of the randomized measurement toolbox for quantum simulators [11–35]. In all of these protocols, a measurement is performed after a random unitary gate, and the desired property is predicted through a classical computer after collecting sufficient measurement outcomes. In particular, the random measurement has been experimentally realized in [36–41]. These developments unveil that randomness plays a central role in extracting information from complex quantum systems efficiently. From a machine learning perspective, this implies that introducing random unitaries can enhance the expressivity of QNNs. This naturally leads to the concept of randomized quantum neural networks, where we collect measurement outcomes from an ensemble of parametrized quantum circuits to make final predictions. Analogous to the different types of layers in classical neural networks, randomized QNNs consist of deterministic layers and random layers. In deterministic layers, the quantum gates contain parametrized quantum gates as in traditional QNNs, while in random layers, they are sampled from trainable ensembles

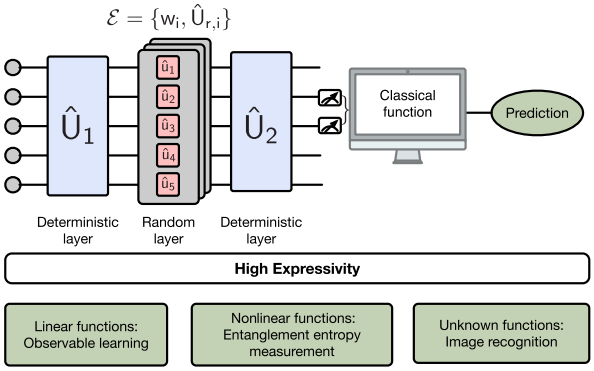

FIG. 1. An illustration is provided for the proposed architecture of randomized quantum neural networks. In this example, the circuit contains two deterministic layers $\hat{U}_{1(2)}$ and one random layer $\hat{U}_r$ in between, with the final measurement performed on two qubits. As demonstrated in this Letter, this architecture shows randomness-enhanced expressivity for a variety of general learning tasks.

of single-qubit gates. This is illustrated in Fig. 1. We demonstrate the high expressivity of the proposed architecture using several different tasks, including both linear and nonlinear functions of the input density matrix. Our results pave the way toward realizing the universal expressivity ability for QNNs.

Architecture.—We begin with a detailed description of randomized QNNs. To be concrete, we focus on the architecture illustrated in Fig. 1 for $N_{\rm sys} = 5$ qubits, which comprises two deterministic layers, namely, $\hat{U}_1$ and $\hat{U}_2$, with a random single-qubit gate layer $\hat{U}_r$ in between.

Each deterministic layer $\hat{U}_{l_d}$ $(l_d = 1, 2)$ contains a number of units $\hat{V}^l_{l_d}(\boldsymbol{\theta}^l_{l_d})$ $(l \in \{1, 2, \ldots, L_{l_d}\})$, and each deterministic layer is constructed as

$$\hat{U}_{l_d} = \hat{V}^{L_{l_d}}_{l_d}(\boldsymbol{\theta}^{L_{l_d}}_{l_d}), \ldots, \hat{V}^2_i(\boldsymbol{\theta}^2_{l_d})\hat{V}^1_{l_d}(\boldsymbol{\theta}^1_{l_d}), \tag{1}$$

where $\{\boldsymbol{\theta}^l_{l_d}\}$ are the parameters of the deterministic layers. In general, the arrangement of two-qubit gates in each deterministic layer allows for a large degree of freedom. In this Letter, we focus on the standard brick wall architecture with spatial locality. Each unit $\hat{V}^l_{l_d}$ contains $N_{\rm sys} - 1$ two qubit gates and each two qubit gate is an SU(4) matrix which can be parametrized as $\exp(\sum_j c_j \hat{g}_j)$. Here, $\hat{g}_j$ is the generator of the SU(4) group, and $\{\boldsymbol{\theta}^l_{l_d}\}$ denotes parameters $\{\mathbf{c}\}$ of all two qubit gates [42]. Nonetheless, alternative choices for each deterministic layer have the potential to enhance the expressivity of QNNs for a fixed number of gates [6].

For the sake of experimental convenience, the random layer $\hat{U}_r$ comprises a tensor product of single-qubit gates denoted as $\hat{u}_1 \otimes \hat{u}_2, \ldots, \otimes \hat{u}_{N_{\rm sys}}$. These gates are sampled from an ensemble

$$\mathcal{E} = \{[w_i, \hat{U}_{r,i} = \hat{u}^i_1(\boldsymbol{\alpha}^1_i) \otimes \hat{u}^i_2(\boldsymbol{\alpha}^2_i), \ldots, \otimes \hat{u}^i_{N_{\rm sys}}(\boldsymbol{\alpha}^{N_{\rm sys}}_i)]\}, \tag{2}$$

where $i = 1, 2, \ldots, N_r$ labels different elements, and w_i is the corresponding weight with $\sum_i w_i = 1$. Each single-qubit gate is parametrized by generators of SU(2) with three-dimensional real vector $\boldsymbol{\alpha}^q_i$ $(q \in \{1, 2, \ldots, N_{\rm sys}\})$. Both $\{w_i\}$ and $\{\boldsymbol{\alpha}^q_i\}$ are trainable parameters. It is also straightforward to introduce multiple random layers into the full architecture of QNNs. Importantly, it is worth noting the differences between our definition and typical random measurement protocols. First, our random layer can be added at any point in the quantum circuit, not necessarily before the final measurement. Second, our definition of $\mathcal{E}$ allows for nontrivial correlations between single-qubit gates on different sites, which is typically absent in random measurement protocols. Both features are necessary for achieving a high expressivity in QNNs.

We consider a dataset $\{(|\psi_m\rangle, \mathcal{T}_m)\}$ in which $m \in \{1, 2, \ldots, N_D\}$ labels different data, and $\mathcal{T}_m$ is the target information for the corresponding state $|\psi_m\rangle$. For each unitary $\hat{U}_{r,i}$ in the ensemble $\mathcal{E}$, we perform projective measurements in the computational basis for $k \sim O(1)$ qubits. The small number of measured qubits would avoid the barren plateaus, which can be caused by global measurements [43]. In Fig. 1, we set $k = 2$, and the measurement yields the probability distribution given by

$$p^{ss'}_{i,m} = \langle\psi_m|\hat{U}^\dagger_1\hat{U}^\dagger_{r,i}\hat{U}^\dagger_2(\hat{P}^2_s \otimes \hat{P}^3_{s'})\hat{U}_2\hat{U}_{r,i}\hat{U}_1|\psi_m\rangle, \tag{3}$$

where the projection operator $\widehat{P^q_s} = (1 + s\hat{\sigma}^q_z)/2$ for $s = \pm 1$. Because of the constraint $\sum_{ss'} p^{ss'}_{i,m} = 1$, there are only three nontrivial components of $p^{ss'}_{i,m}$ denoted by the vector $\mathbf{p}_{i,m}$. We then use a classical computer to apply a general function $f_{\boldsymbol{\beta}}(.)$ parametrized by $\boldsymbol{\beta}$, to the probability distribution $p^{ss'}_{i,m}$, which yields a single outcome denoted by $\mathcal{P}_{i,m} = f_{\boldsymbol{\beta}}(\mathbf{p}_{i,m})$. The classical function can be described by elementary functions in the simplest setting, but is more generally described by classical neural networks. We further average the outcome over the ensemble $\mathcal{E}$ to obtain the final prediction for the input state $|\psi_m\rangle$ as

$$\mathcal{P}_m = \sum_{i=1}^{N_r} w_i \mathcal{P}_{i,m} = \sum_{i=1}^{N_r} w_i f_{\boldsymbol{\beta}}(\mathbf{p}_{i,m}). \tag{4}$$

We use the mean square error (MSE) as the loss function $\mathcal{L} = (1/N_D)\sum_m (\mathcal{P}_m - \mathcal{T}_m)^2$ with a data size of N_D during the training process. We apply the gradient descent algorithm to optimize the parameters $\{\boldsymbol{\theta}^l_{l_d}, w_i, \boldsymbol{\alpha}^q_i, \boldsymbol{\beta}\}$ to minimize the loss function $\mathcal{L}$, and set the numerical criteria as $\mathcal{L} < 10^{-5}$ to characterize the accurate prediction. Our method to compute gradients of parameters is explained in the Supplemental Material [42]. In the following sections, we focus on demonstrating high expressivity for randomized QNNs. Our examples range from simple physical tasks including observable learning and Rényi entropy

measurement, to standard machine learning tasks such as image recognition.

Observable learning.—To show the high expressivity of randomized QNNs, let us consider a simple scenario where the target $\mathcal{T}_m$ is an expectation of a physical observable $\hat{O}$ with $\mathcal{T}_m = \langle\psi_m|\hat{O}|\psi_m\rangle$. For simplicity, focusing on single-qubit measurement with $k = 1$, we first investigate whether the randomized QNNs as proposed in Fig. 1 can approximate the target function $\mathcal{T}_m$ as accurately as possible for sufficiently deep circuit structures with sufficiently large N_r. As physical observables are linear in density matrices, a linear function $f_{\boldsymbol{\beta}}(x) = \beta_0 + \beta_1 x$ will be applied to the measurement result. Explicitly, we introduce $\hat{U}_{\mathrm{tot},i}$ for a random realization i of the quantum circuit. As an example, we have $\hat{U}_{\mathrm{tot},i} = \hat{U}_2\hat{U}_{r,i}\hat{U}_1$. An accurate prediction of the target function requires that

$$\sum_{i=1}^{N_r} w_i \hat{U}_{\mathrm{tot},i}^{\dagger}(\beta_0\hat{\sigma}_0^1 + \beta_1\hat{\sigma}_z^1)\hat{U}_{\mathrm{tot},i} = \hat{O}, \tag{5}$$

where $\hat{\sigma}_0$ is the identity operator, and Pauli matrix $\hat{\sigma}_z$ is the single qubit's measurement operator.

For the case of $N_r = 1$ and $w_1 = 1$, our setup reduces to the traditional QNN without randomness. In this scenario, Eq. (5) requires that $(\beta_0\hat{\sigma}_0^1 + \beta_1\hat{\sigma}_z^1)$ and $\hat{O}$ be related by a unitary transformation. Since the unitary transformation preserves the eigenvalues of the operator, the requirement cannot be satisfied for a general operator $\hat{O}$. When $N_r > 1$, Eq. (5) can be expressed as $\Phi(\hat{\Sigma}) = \hat{O}$, where $\hat{\Sigma} \equiv \beta_1\hat{\sigma}_z^1 + \beta_0\hat{\sigma}_0^1$ and $\Phi(\hat{X})$ is a mixed-unitary channel [44]. For sufficiently complex circuit structures, we expect Φ to be generic. In comparison to the $N_r = 1$ case, there is no constraint from unitarity. However, we still need to ask whether Eq. (5) can be satisfied for an arbitrary operator $\hat{O}$. In the following, we prove that the answer to this question is affirmative.

Step 1: Mathematically, if there exists a mixed-unitary channel Φ such that $Y = \Phi(X)$, we say that X majorizes Y, denoted by $Y \prec X$ [45]. Thus, for randomized QNNs, which can accurately predict any observable $\hat{O}$, we need to find values of β_0 and β_1 such that $\hat{O} \prec \hat{\Sigma}$ for any $\hat{O}$.

Step 2: According to Uhlmann's theorem for majorization [45,46], $\hat{O} \prec \hat{\Sigma}$ if and only if $\boldsymbol{\lambda}_{\hat{O}} \prec \boldsymbol{\lambda}_{\hat{\Sigma}}$, where $\boldsymbol{\lambda}_{\hat{X}}$ is the list of eigenvalues for the operator $\hat{X}$ in descending order. Here the majorization between two real vectors $\boldsymbol{y} \prec \boldsymbol{x}$ is defined as (i) $\sum_{j=1}^{q} x_j \geq \sum_{j=1}^{q} y_j$ for arbitrary $1 \leq q < \mathcal{D}$ and (ii) $\sum_{j=1}^{\mathcal{D}} x_j = \sum_{j=1}^{\mathcal{D}} y_j$. Here, $\mathcal{D}$ is the dimension of the vectors. We note that condition (ii) takes into account the trace-preserving property of mixed-unitary channels.

Step 3: We can always find β_0 and β_1 such that $\boldsymbol{\lambda}_{\hat{O}} \prec \boldsymbol{\lambda}_{\hat{\Sigma}}$. Assuming $\beta_1 > 0$, the first or last $\mathcal{D}/2$ components of $\boldsymbol{\lambda}_{\hat{\Sigma}}$ correspond to the values $\beta_0 + \beta_1$ or $\beta_0 - \beta_1$, respectively. The constant term β_0 can then be determined using condition (ii), which gives $\beta_0 = \mathcal{D}^{-1}\sum_{j=1}^{\mathcal{D}} \lambda_{\hat{O},j}$. Moreover, condition (i) can always be satisfied for sufficiently large β_1. This proves the existence of β_0 and β_1 such that $\hat{O} \prec \hat{\Sigma}$.

Although randomized QNNs have the potential to express arbitrary operators, it is difficult to determine an upper bound or a required value for N_r in practical learning tasks. It is unfavorable to have large N_r or a large number of random layers, especially in NISQ devices. Therefore, we turn to numerical simulations of the randomized QNNs, and investigate practical requirements on N_r. Since the basis change can be efficiently captured by the deterministic layer $\hat{U}_1$, we focus on observables $\hat{O}$ that are diagonal in the computational basis. For simplicity, we further set $\hat{U}_1 = \hat{I}$ and $\hat{U}_2$ composed by L_2 units of a brick wall structure [42]. For each system size N_{sys}, we test whether a random diagonal operator $\hat{O}$ can be predicted accurately for different values of N_r by monitoring the training loss for a sufficiently large dataset. As an example, we plot the logarithmic training MSE $\log_{10}$ as a function of the training epoch for $N_{\mathrm{sys}} = 2$ in Fig. 2(a). The curves are averaged

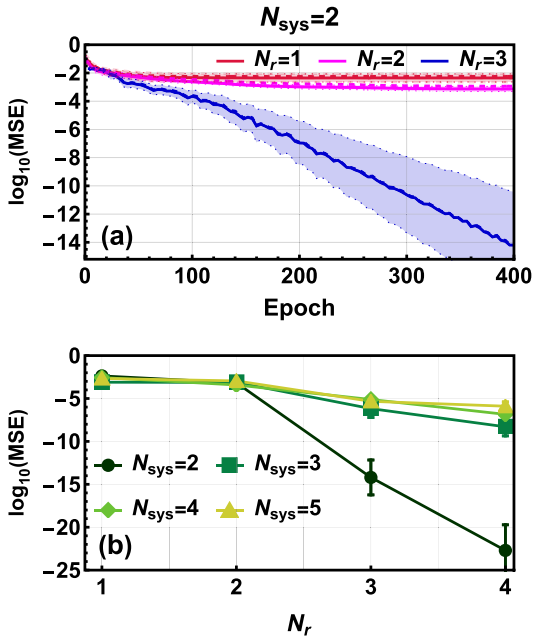

FIG. 2. Predicting observables using QNN with a random layer. (a) The logarithmic training mean square error is shown as a function of the training epoch for observable learning with $N_{\mathrm{sys}} = 2$. The solid lines represent the averaged over the training process for ten different random target operators with independent runs, while the shaded region represents the standard deviation. The dashed lines are the validation loss with the dataset containing 200 samples. (b) The logarithmic mean square error of training dataset for $N_{\mathrm{sys}} \in \{2, 3, 4, 5\}$ and $N_r \in \{1, 2, 3, 4\}$. The markers represents the average over ten different random target operators with random initializations, and error bars are the standard deviation.

over ten operators with random eigenvalues from the uniform distribution $[-2.5, 2.5]$. When we increase N_r from 1 to 3, there is a rapid decrease in the training loss for large training epochs. The result shows that $N_r = 3$ is sufficient for learning general operators for $N_{\mathrm{sys}} = 2$ where the loss $\mathcal{L}$ can be decreased to 10^{-14}. We further extend the system size N_{sys} to study how it affects the number of required random gates. The results are shown in Fig. 2(b). Although we are limited to small system sizes $N_{\mathrm{sys}} \in \{2, 3, 4, 5\}$, the results clearly show weak dependence of N_r on N_{sys}. The training results show that $N_r = 3$ already gives highly accurate predictions for $N_{\mathrm{sys}} = 5$.

Rényi entropy measurement.—We now consider targets that are nonlinear functions of density matrices. One example is the Rényi entropy, which is also of experimental interest. To compute the Rényi entropy for a subsystem A consisting of the central N_{sub} qubits, we first calculate the reduced density matrix $\hat{\rho}_A$ of an input state $|\psi_m\rangle$ by tracing out the degrees of freedom of the complementary subsystem $\overline{A}$. We then select the target as

$$\mathcal{T}_m = \mathrm{Tr}_A[\hat{\rho}_A^n], \tag{6}$$

which is related to the nth Rényi entropy through $S_A^{(n)} = -[1/(n-1)]\ln(\mathcal{T}_m)$. Since we are directly measuring a local property of the input wave function, it is reasonable to fix $\hat{U}_1$ and $\hat{U}_2$ to the identity matrix $\hat{I}$ and focus on the random layer $\hat{U}_r$ with $k = N_{\mathrm{sub}}$. This approach provides a minimum guaranteed expressivity of randomized QNNs. Because the target $\mathcal{T}_m$ is proportional to ρ^n, we choose the function $f_{\boldsymbol{\beta}}(\mathbf{x})$ to be a polynomial up to the nth order. However, it is worth noting that lower-order polynomials may also work in certain cases [47]. We prepare a dataset with random states $|\psi_m\rangle$, the detailed description of which is provided in the Supplemental Material [42]. The numerical results for $n = 2, 3$, $N_{\mathrm{sys}} = 5$, and $N_{\mathrm{sub}} = 1, 2$ are shown in Fig. 3. To achieve accurate predictions, we need $N_r = 3$ for $N_{\mathrm{sub}} = 1$ and $N_r = 9$ for $N_{\mathrm{sub}} = 2$. The blue lines in Fig. 3 demonstrate that the loss $\mathcal{L}$ is able to reach a value of 10^{-5} and still keep decreasing, indicating the ability to make accurate predictions. We have also discussed the saturation of N_r for $n = 2, N_{\mathrm{sub}} = 2$, and the required number of N_r if we instead consider $n = 3$. The results are shown in the Supplemental Material [42].

It is interesting to compare our results to the proposed random measurement protocol for Rényi entropies. Our results indicate that N_r scales as $3^{N_{\mathrm{sub}}}$ when measuring Rényi entropies. In comparison, the previous protocol required each single-qubit gate $\hat{u}_q^i$ to be sampled from the circular unitary ensemble [18,40]. For $n = 2$, the circular unitary ensemble can be replaced by unitary 2-designs, which are known to be the Clifford group. Since the single-qubit Clifford group contains 24 elements, the total number of unitary matrices $\hat{U}_{r,i}$ would naively scale as $24^{N\mathrm{sub}}$. However, in practice, this can be significantly reduced because randomized measurement protocols only require N_s snapshots sampled from the full ensemble. The theoretical bound of N_s for measuring general linear observables in a subsystem with N_{sub} qubits using random Pauli measurements up to an error ϵ is given by $N_s \gtrsim 3^{N_{\mathrm{sub}}}/\epsilon^2$ [18,19]. Consequently, in this quantum neural network structure, the number of unitary matrices $\hat{U}_{r,i}$ that contribute is at most $3^{N_{\mathrm{sub}}}$, as in our randomized QNNs.

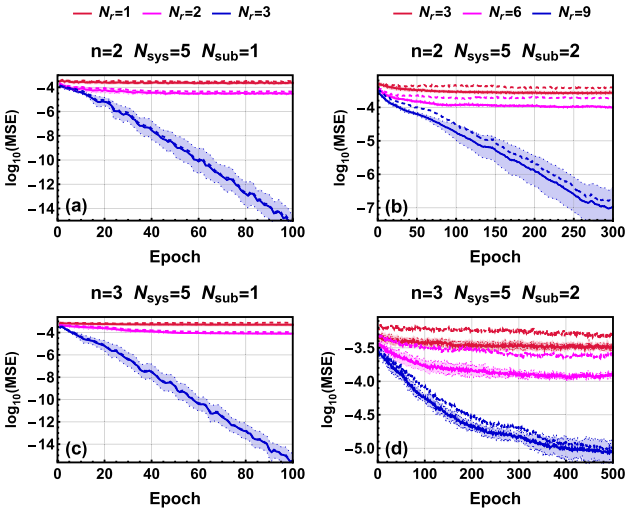

FIG. 3. Predicting Rényi entropies using QNN with a random layer. The logarithmic training mean square error is shown as a function of the training epoch for purity with $N_{\mathrm{sys}} = 5$ and (a) $n = 2, N_{\mathrm{sub}} = 1$, (b) $n = 2, N_{\mathrm{sub}} = 2$, (c) $n = 3, N_{\mathrm{sub}} = 1$, or (d) $n = 3, N_{\mathrm{sub}} = 2$. The results are averaged over the training process for ten different random initializations, and the shaded region represents the standard deviation. The dashed lines are the validation loss with the dataset containing 200 samples.

Image recognition.—Finally, we turn our attention to image recognition, a more practical machine learning task, in order to demonstrate the enhanced expressivity of randomized QNNs. In this case, we use Google's "Street View of House Number" dataset as an example [48]. Each image in the dataset corresponds to an integer number. For demonstration purposes, we select two categories of images containing the numbers "1" and "4." Initially, we compress each image into an 8×8 pixel format, resulting in a 64-dimensional real vector, which can be equivalently represented as a 32-dimensional complex vector. Subsequently, we encode the image into the input wave function using $N_{\mathrm{sys}} = 5$ qubits [42]. Unlike previous tasks, the mapping between the input and the output is highly complex and nonlocal, lacking a simple understanding. Consequently, we allow both $\hat{U}_1$ and $\hat{U}_2$ to be trainable. After measuring a single qubit, we choose a fifth-order polynomial for the function $f_{\boldsymbol{\beta}}(\boldsymbol{x})$. Since the image recognition is a two-category classification task, after obtaining the final ensemble average prediction $\mathcal{P}_m$, we apply a logistic-sigmoid

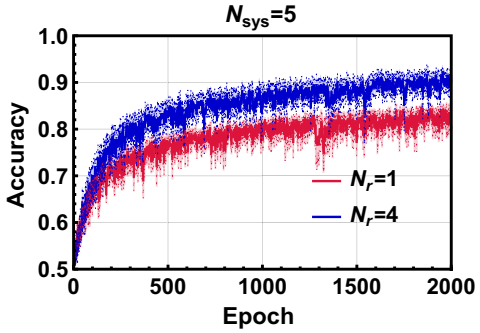

FIG. 4. Image recognition by QNN with a random layer. The accuracy is shown as a function of the training epoch for the image recognition task. The results are averaged over the training process for ten different random initializations, and the shaded region represents the standard deviation.

function to restrict the prediction in the region (0,1) with $\mathcal{G}_m = 1/[1+\exp(-\mathcal{P}_m)]$, and we use the cross-entropy as the loss function $\mathcal{L} = (1/N_D)\sum_m -\mathcal{T}_m \log(\mathcal{G}_m) - (1-\mathcal{T}_m)\log(1-\mathcal{G}_m)$ to optimize the parameters in the randomized QNN. The accuracy $F = (1/N_D)\sum_m |[\mathrm{sign}(\mathcal{G}_m - 0.5)+1]/2 - \mathcal{T}_m|$ for $N_r = 1$ and $N_r = 4$ is shown in Fig. 4. For $N_r = 1$, the accuracy saturates at approximately 0.8 after a large number of epochs, while the averaged accuracy for the test dataset reaches 69.8%. The introduction of a single random layer with $N_r = 4$ significantly enhances the accuracy of the predictions. In this case, the training dataset achieves an accuracy higher than 90%, and the average accuracy for the test dataset is 82.29%. The utilization of a nontrivial random layer with $N_r = 4$ demonstrates a significant improvement in the prediction capabilities of QNNs, indicating the enhanced expressivity of our randomized QNN architecture.

Outlook.—This Letter introduces the concept of randomized quantum neural networks, which include random layers where quantum gates are selected from an ensemble of unitary matrices. It is proven that these random layers provide universal expressivity for general physical observables using Uhlmann's theorem for majorization. Numerical simulations further show that this architecture achieves high expressivity for nonlinear functions of the density matrix, such as Rényi entropies and image recognition, with small ensemble sizes N_r. These results indicate that the proposed method has potential for broad applications in NISQ devices. We remark that adding a random layer to the QNNs causes an extra computational cost proportional to N_r. Nonetheless, introducing randomness to QNNs while maintaining the same computational cost still improves the learning performance significantly [42]. We further highlight the differences between our architecture and the proposal presented in a very recent paper [49]. Their work also incorporates a series of parametrized quantum circuits, where the circuit consists of multiple parametrized (controlled) rotations that share the same parameter. In contrast, our architecture features only a few random layers described by a tensor product of single-qubit gates, making its training process more efficient.

While the focus of this Letter is on parametrized quantum circuits with brick wall structures, it is straightforward to combine this novel architecture with other proposals to further improve expressivity or learning efficiency. For instance, it is possible to add ancilla qubits and explore more sophisticated architectures for the deterministic layers. Additionally, it would be interesting to investigate the impact of random layers on other quantum machine learning algorithms beyond traditional quantum neural networks [50–55], such as quantum autoencoders [50,51].

We thank Yingfei Gu, Ning Sun, Ce Wang, Hai Wang, and Yi-Zhuang You for helpful discussions. This work is supported by National Program on Key Basic Research Project of China (Grant No. 2021YFA1400900), National Natural Science Foundation of China (Grants No. 11934002, No. 12174236, No. 11904190), and Shanghai Municipal Science and Technology Major Project (Grant No. 2019SHZDZX01).

*pengfeizhang.physics@gmail.com
†xiaopeng_li@fudan.edu.cn

[1] Maria Schuld, Alex Bocharov, Krysta M. Svore, and Nathan Wiebe, Circuit-centric quantum classifiers, Phys. Rev. A **101**, 032308 (2020).
[2] Jin-Guo Liu and Lei Wang, Differentiable learning of quantum circuit born machines, Phys. Rev. A **98**, 062324 (2018).
[3] Edward Farhi and Hartmut Neven, Classification with quantum neural networks on near term processors, arXiv:1802.06002.
[4] Marcello Benedetti, Erika Lloyd, Stefan Sack, and Mattia Fiorentini, Parameterized quantum circuits as machine learning models, Quantum Sci. Technol. **4**, 043001 (2019).
[5] Yadong Wu, Juan Yao, Pengfei Zhang, and Hui Zhai, Expressivity of quantum neural networks, Phys. Rev. Res. **3**, L032049 (2021).
[6] Yadong Wu, Pengfei Zhang, and Hui Zhai, Scrambling ability of quantum neural network architectures, Phys. Rev. Res. **3**, L032057 (2021).
[7] Leonardo Banchi, Jason Pereira, and Stefano Pirandola, Generalization in quantum machine learning: A quantum information standpoint, PRX Quantum **2**, 040321 (2021).
[8] Matthias C. Caro, Hsin-Yuan Huang, M. Cerezo, Kunal Sharma, Andrew Sornborger, Lukasz Cincio, and Patrick J. Coles, Generalization in quantum machine learning from few training data, Nat. Commun. **13**, 4919 (2022).
[9] Amira Abbas, David Sutter, Christa Zoufal, Aurelien Lucchi, Alessio Figalli, and Stefan Woerner, The power of quantum neural networks, Nat. Comput. Sci. **1**, 403 (2021).
[10] Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer, Effect of data encoding on the expressive power of

variational quantum-machine-learning models, Phys. Rev. A **103**, 032430 (2021).
[11] Xiao-Liang Qi, Emily J. Davis, Avikar Periwal, and Monika Schleier-Smith, Measuring operator size growth in quantum quench experiments, arXiv:1906.00524.
[12] Scott Aaronson, Shadow tomography of quantum states, arXiv:1711.01053.
[13] Andreas Ketterer, Nikolai Wyderka, and Otfried Gühne, Characterizing multipartite entanglement with moments of random correlations, Phys. Rev. Lett. **122**, 120505 (2019).
[14] A. Elben, B. Vermersch, M. Dalmonte, J. I. Cirac, and P. Zoller, Rényi entropies from random quenches in atomic Hubbard and spin models, Phys. Rev. Lett. **120**, 050406 (2018).
[15] S. J. van Enk and C. W. J. Beenakker, Measuring $\mathrm{Tr}\rho^n$ on single copies of ρ using random measurements, Phys. Rev. Lett. **108**, 110503 (2012).
[16] A. Elben, B. Vermersch, C. F. Roos, and P. Zoller, Statistical correlations between locally randomized measurements: A toolbox for probing entanglement in many-body quantum states, Phys. Rev. A **99**, 052323 (2019).
[17] Lukas Knips, Jan Dziewior, Waldemar Kłobus, Wiesław Laskowski, Tomasz Paterek, Peter J. Shadbolt, Harald Weinfurter, and Jasmin D. A. Meinecke, Multipartite entanglement analysis from random correlations, npj Quantum Inf. **6**, 51 (2020).
[18] Hsin-Yuan Huang, Richard Kueng, and John Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16**, 1050 (2020).
[19] Andreas Elben, Steven T. Flammia, Hsin-Yuan Huang, Richard Kueng, John Preskill, Benoît Vermersch, and Peter Zoller, The randomized measurement toolbox, Nat. Rev. Phys. **5**, 9 (2023).
[20] Hong-Ye Hu and Yi-Zhuang You, Hamiltonian-driven shadow tomography of quantum states, Phys. Rev. Res. **4**, 013054 (2022).
[21] Senrui Chen, Wenjun Yu, Pei Zeng, and Steven T. Flammia, Robust shadow estimation, PRX Quantum **2**, 030348 (2021).
[22] Atithi Acharya, Siddhartha Saha, and Anirvan M. Sengupta, Informationally complete POVM-based shadow tomography, arXiv:2105.05992.
[23] Ryan Levy, Di Luo, and Bryan K. Clark, Classical shadows for quantum process tomography on near-term quantum computers, arXiv:2110.02965 [Phys. Rev. Res. (to be published)], https://journals.aps.org/prresearch/accepted/8207eK39Y591580ed6d94007613fe8001ae2078ad.
[24] Andrew Zhao, Nicholas C. Rubin, and Akimasa Miyake, Fermionic partial tomography via classical shadows, Phys. Rev. Lett. **127**, 110504 (2021).
[25] Kianna Wan, William J. Huggins, Joonho Lee, and Ryan Babbush, Matchgate shadows for fermionic quantum simulation, Commun. Math. Phys. **404**, 629 (2023).
[26] Hsin-Yuan Huang, Richard Kueng, Giacomo Torlai, Victor V. Albert, and John Preskill, Provably efficient machine learning for quantum many-body problems, Science **377**, eabk3333 (2022).
[27] Kaifeng Bu, Dax Enshan Koh, Roy J. Garcia, and Arthur Jaffe, Classical shadows with Pauli-invariant unitary ensembles, arXiv:2202.03272.
[28] Jonathan Kunjummen, Minh C. Tran, Daniel Carney, and Jacob M. Taylor, Shadow process tomography of quantum channels, Phys. Rev. A **107**, 042403 (2023).
[29] Saumya Shivam, Curt W. von Keyserlingk, and Shivaji L. Sondhi, On classical and hybrid shadows of quantum states, SciPost Phys. **14**, 094 (2023).
[30] Matteo Ippoliti, Yaodong Li, Tibor Rakovszky, and Vedika Khemani, Operator relaxation and the optimal depth of classical shadows, Phys. Rev. Lett. **130**, 230403 (2023).
[31] Christian Bertoni, Jonas Haferkamp, Marcel Hinsche, Marios Ioannou, Jens Eisert, and Hakop Pashayan, Shallow shadows: Expectation estimation using low-depth random Clifford circuits, arXiv:2209.12924.
[32] Mirko Arienzo, Markus Heinrich, Ingo Roth, and Martin Kliesch, Closed-form analytic expressions for shadow estimation with brickwork circuits, Quantum Inf. Comput. **23**, 961 (2023).
[33] Ahmed A. Akhtar, Hong-Ye Hu, and Yi-Zhuang You, Scalable and flexible classical shadow tomography with tensor networks, Quantum **7**, 1026 (2023).
[34] Hong-Ye Hu, Soonwon Choi, and Yi-Zhuang You, Classical shadow tomography with locally scrambled quantum dynamics, Phys. Rev. Res. **5**, 023027 (2023).
[35] Dax Enshan Koh and Sabee Grewal, Classical shadows with noise, Quantum **6**, 776 (2022).
[36] Joseph Vovrosh and Johannes Knolle, Confinement and entanglement dynamics on a digital quantum computer, Sci. Rep. **11**, 11577 (2021).
[37] Min Yu, Dongxiao Li, Jingcheng Wang, Yaoming Chu, Pengcheng Yang, Musang Gong, Nathan Goldman, and Jianming Cai, Experimental estimation of the quantum Fisher information from randomized measurements, Phys. Rev. Res. **3**, 043122 (2021).
[38] Crystal Noel, Pradeep Niroula, Daiwei Zhu, Andrew Risinger, Laird Egan, Debopriyo Biswas, Marko Cetina, Alexey V. Gorshkov, Michael J. Gullans, David A. Huse, and Christopher Monroe, Measurement-induced quantum phases realized in a trapped-ion quantum computer, Nat. Phys. **18**, 760 (2022).
[39] Jin Ming Koh, Shi-Ning Sun, Mario Motta, and Austin J. Minnich, Measurement-induced entanglement phase transition on a superconducting quantum processor with mid-circuit readout, Nat. Phys. **19**, 1314 (2023).
[40] Tiff Brydges, Andreas Elben, Petar Jurcevic, Benoît Vermersch, Christine Maier, Ben P. Lanyon, Peter Zoller, Rainer Blatt, and Christian F. Roos, Probing Renyi entanglement entropy via randomized measurements, Science **364**, 260 (2019).
[41] G. I. Struchalin, Ya. A. Zagorovskii, E. V. Kovlakov, S. S. Straupe, and S. P. Kulik, Experimental estimation of quantum state properties from classical shadows, PRX Quantum **2**, 010307 (2021).
[42] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.132.010602 for (i) structures of deterministic layers, (ii) gradient decent method of QNN's parameters, (iii) training details of the observables learning task, (iv) training details of the Rényi entropy measurement task, (v) training details of the pattern recognition task, and (vi) comparison with fixed computational cost.

[43] Marco Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, Nat. Commun. **12**, 1791 (2021).

[44] John Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, England, 2018), https://doi.org/10.1017/9781316848142.

[45] Michael A. Nielsen, *An Introduction to Majorization and Its Applications to Quantum Mechanics*, Lecture Notes, Department of Physics (University of Queensland, Australia, 2002).

[46] Peter Alberti and Armin Uhlmann, *Stochasticity and Partial Order. Doubly Stochastic Maps and Unitary Mixing* (Springer, Dordrecht, 1982), https://link.springer.com/book/9789027713506.

[47] The third order of single-qubit reduced density matrix $\mathrm{tr}[\hat{\rho}_a^3]$ is the second order of the elements of the density matrix.

[48] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y. Ng, Reading digits in natural images with unsupervised feature learning NIPS workshop on deep learning and unsupervised feature learning (2011), http://ufldl.stanford.edu/housenumbers/.

[49] Xiaokai Hou, Guanyu Zhou, Qingyu Li, Shan Jin, and Xiaoting Wang, A duplication-free quantum neural network for universal approximation, Sci. China Phys. Mech. Astron. **66**, 270362 (2023).

[50] Jonathan Romero, Jonathan P Olson, and Alan Aspuru-Guzik, Quantum autoencoders for efficient compression of quantum data, Quantum Sci. Technol. **2**, 045001 (2017).

[51] Dmytro Bondarenko and Polina Feldmann, Quantum autoencoders to denoise quantum data, Phys. Rev. Lett. **124**, 130502 (2020).

[52] Seunghyeok Oh, Jaeho Choi, and Joongheon Kim, A tutorial on quantum convolutional neural networks (QCNN), in *Proceedings of the 2020 International Conference on Information and Communication Technology Convergence (ICTC)* (IEEE, Jeju, Korea (South), 2020), pp. 236–239, https://dx.doi.org/10.1109/ICTC49870.2020.9289439.

[53] Tak Hur, Leeseok Kim, and Daniel K. Park, Quantum convolutional neural network for classical data classification, Quantum Mach. Intell. **4**, 3 (2022).

[54] Iris Cong, Soonwon Choi, and Mikhail D. Lukin, Quantum convolutional neural networks, Nat. Phys. **15**, 1273 (2019).

[55] Matthias C. Caro, Hsin-Yuan Huang, M. Cerezo, Kunal Sharma, Andrew Sornborger, Lukasz Cincio, and Patrick J. Coles, Generalization in quantum machine learning from few training data, Nat. Commun. **13**, 4919 (2022).